

## イベント-アクションの概念に基づいた動画像のシーン検索手法\*

5 G-6

牛尾剛聡† 広部一弥† 渡邊豊英†  
名古屋大学大学院工学研究科情報工学専攻‡

### 1 はじめに

動画像は対象世界の動的な側面を記録するのに適した表現形式である。たとえば、野球の試合における投手の投球フォームや打球の軌跡のような現象の様子は、動画像を用いれば人間にとって直観的にわかりやすく表現可能である。しかし、動画像中には様々な現象が連続的、並行的に発生しているため、動画像中から利用者の興味のある現象に関する部分だけを提供する機構が要求されている。

シーンとは一つの現象を表現する動画像中の部分である。対象世界の現象を表現するための概念は多種多様であり、検索を行うユーザの視点によって、対象とする現象の捉え方は異なる。我々は、ユーザの様々な視点に基づいた動画像のシーン検索を可能とする動画像情報システム STRIKE(Stream data Retrieval based on Indexing with Key Event)を開発中である。本稿では STRIKE における基本的なシーン検索の枠組について述べる。

### 2 アプローチ

我々の基本的な考え方として、動的な現象を出来事の系列として定義する。たとえば、野球の試合において、「投手が投球する」、「打者が球を打つ」、「打球が直接外野スタンドに入る」という出来事が順に発生したとする。このとき、人はこれらの一連の出来事を「ホームランを打つ」という現象として認識する。そこで、動画像中の各フレームに対して発生した出来事を表すインデックスを付加しておき、定義された出来事の系列を含むような部分を捜し出すことにより、シーンを検索することができる。上記の例における検索の概念図を図1に示す。

これまでにも、シーン検索に関して、いくつかの手法が提案されている。それらの手法の多くは、検索対象であるシーンを単位としてインデックスを付加している。一方、我々の手法は、フレームを単位としてインデックスを付加し、ユーザの要求に従って、フレーム順に基づくインデックスの系列パターンから動的にシーンを構成する。前者は現象を個別的な「モノ」として捉えるアプローチであり、後者は現象を連続的な「コト」として捉えるアプローチである。対象世界上の現象は多重的、複合的に発生しているため、物体のように各々を排他的

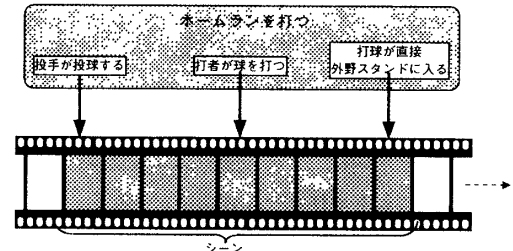


図1: 現象「ホームランを打つ」を表すシーンの検索

に個別化することは冗長かつ困難である。したがって、現象を連続的な「コト」として捉えるアプローチのほうが、様々な視点からの現象の概念を扱うのに適していると考えられる。

### 3 イベント-アクション・モデル

上記のアプローチに基づいた動画像のインデックス付けとシーン検索を実現するイベント-アクション・モデルを提案する。イベント-アクション・モデルは、動画像中の変化を表すイベントと、イベントの系列パターンの定義規則であるアクションにより構成される。イベントは動画像中の出来事を表し、アクションは動画像中の出来事の並びを現象として解釈するための知識である。

#### 3.1 状態と状況

動画像は、フレーム(2次元静止画像)の系列である。フレームは対象世界上の物体のスナップショットを保持している。フレーム上で、各々の物体のスナップショットを表現している閉領域を状態と呼ぶ。それぞれのフレームにおける物体の静的な性質は状態の属性として表現される。状態は、イベント-アクション・モデルにおける原子的な概念である。動画像に含まれる状態の内、同一の物体を表す全ての状態から構成される集合を実体と呼ぶ。同一の実体に属する状態は同一の属性構造を持つものとする。

同一のフレーム内に複数の状態が存在することがある。フレーム  $f_t$  内に存在する状態が属する実体の集合、およびそれらの状態間で成り立つ空間的な関係の集合を  $f_t$  の状況と呼ぶ。状態がフレームの局所的な性質を表現するのに対し、状況はフレーム中の大局的な性質を表現する。状況を考える際に、どのような空間的關係を用いるかは重要である。Liらは動画像における空間的關係として12種類の方向的関係と6種類の幾何学的関係を提案している[1]。しかし、必要となる空間的關係は、対象や応用に依存するため、本稿ではそれらについて詳しく言及しない。図2に状態と状況の例を示す。

\*An Approach for Retrieving Scenes from Movie Based on Event-Action Concept

†Taketoshi USHIAMA, Kazuya HIROBE and Toyohide WATANABE

‡Department of Information Engineering, Graduate School of Engineering, Nagoya University

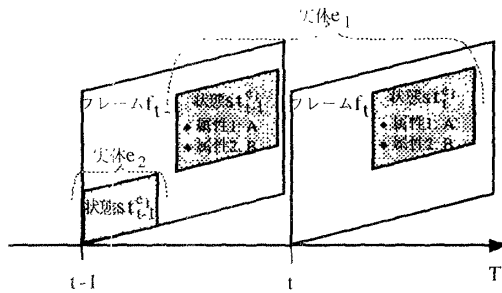


図 2: 状態と状況

### 3.2 イベント

対象世界の出来事を，動画像中の同一実体に属する状態の変化と状況の変化として表すために，イベントを導入する．実体  $e$  に属する連続する 2 状態  $st_{t-1}^e$ ,  $st_t^e$  に対し， $st_{t-1}^e$  と  $st_t^e$  の属性値の差異を状態イベントと呼ぶ．ここで，状態  $st_{t-1}^e$ ,  $st_t^e$  が連続するとは，それぞれが連続するフレーム  $f_{t-1}$ ,  $f_t$  に存在することである．一方， $f_{t-1}$  の状況  $st_{t-1}$  と  $f_t$  の状況  $st_t$  の差異を状況イベントと呼ぶ．状況イベントは以下の 4 種類に分類できる．

1. 実体の出現
2. 実体の消滅
3. 空間的関係の出現
4. 空間的関係の消滅

図 2 に示した例では，実体  $e_1$  の属性 1 の値が  $A$  から  $A'$  になる状態イベントが存在する．また，実体  $e_2$  が消滅する状況イベントが存在する．

### 3.3 イベントの解釈

前節で導入した状態 / 状況イベントは動画像中の物理的な変化を表す．しかし，出来事を表す概念は，そのような物理的な変化を抽象化したものが多い．たとえば，野球の試合における状況イベント「バットとボールの接触」は，一般に出来事を表す概念「打つ」で表現される．このように，イベントを分類し名付けることを解釈と呼ぶ．

野球の試合における「得点」や「アウト・カウント」などのように，動画像中のいずれのフレームにおいても物理的な領域を占有しない概念的なモノが存在する場面がある．概念的なモノも，物体と同様に時間とともに変化するため，概念的なモノに対しても状態や実体を考えることができる．概念的なモノの変化は物理的な変化を解釈することによって生じる．たとえば，ある条件下で，状況イベント「走者が本塁と接触」が発生した場合，概念的な実体「得点」に変化が生じる．

このように，物理イベントを解釈することによって発生するイベントを概念イベントと呼ぶ．また，概念イベントの定義を概念イベント生成知識と呼ぶ．

### 3.4 イベント系列の解釈とアクション

イベントの前後関係によってイベントの解釈が異なる場合がある．たとえば，野球の試合における状況イベン

ト「走者が本塁と接触」に対して，以前に状況イベント「走者とボールを保持した野手との接触」が発生しているとき，前者は「アウト」という概念イベントとして解釈されるが，そうでない場合は概念イベント「ホームイン」として解釈される．これらは概念イベント生成知識に含まれる．

また，2章において，現象「ホームラン」の例で示したように，一つの意味的なまとまりを構成するイベント系列パターンが存在する．現象を表すイベントの系列パターンをアクションと呼ぶ．また，イベントに基づくアクションの定義をアクション生成知識と呼ぶ．

## 4 動画像のインデックス付けとシーン検索

動画像中のフレームにイベント対応付けることで，動画像のインデックス付けを行う．必ずしも，動画像中で観測可能な全てのイベントを対応付ける必要はなく，検索に必要なイベントを選択しインデックスとする．状態や状況を直接インデックスとして記録することはしないが，それぞれのフレーム内の状態や状況はイベント系列から導出可能である．ユーザの問合せは，システムが保持する概念イベント生成知識や，アクション生成知識に基づいて，アクションの定義を与えることによって行う．シーン検索は与えられたアクションに対して，イベント系列の部分系列を捜し出すことによって行なう．

概念イベント生成知識やアクション生成知識は，イベント系列の集合を言語とする形式言語の文法として位置付けられる．このとき，アクションに一致するイベント系列を捜し出す処理は，イベント系列を入力とするオートマトンとして形式化できる．

## 5 おわりに

我々はこれまでに，本稿で示したアプローチに基づく単純な動画像検索機構のプロトタイプを実装し，動作を確認した．実装したプロトタイプでは，概念イベントを手動で動画像にインデックス付けし，アクション生成知識を手続き的に記述している．

本稿で示したイベント・アクション・モデルでは，動画像中で起こるカメラの切替や必要なイベントの省略に対処できない等の問題がある．今後，これらの問題に対処可能なモデルに拡張する．

本稿で示したアプローチに基づいて，「決勝打」などの試合の大局的な文脈に依存した現象を表現することも可能であると考えられる．今後，こうした現象を用いて，動画像のダイジェストを自動生成することを考えている．

## 参考文献

- [1] J. S. Li, T. Özsu and D. Szafron: "Modeling of Video Spatial Relationships in an Object Database Management System", *Proc. of International Workshop Multimedia DBMS*, pp. 124-133 (1996).